HP StorageWorks Continuous Access EVA performance estimator user guide

Part number: T3687–96020
2nd edition: May 2005

Legal and notice information

Copyright © 2005 Hewlett–Packard Development Company, L.P.

Hewlett-Packard Company makes no warranty of any kind with regard to this material, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. Hewlett-Packard shall not be liable for errors contained herein or for incidental or consequential damages in connection with the furnishing, performance, or use of this material.

This document contains proprietary information, which is protected by copyright. No part of this document may be photocopied, reproduced, or translated into another language without the prior written consent of Hewlett-Packard. The information is provided "as is" without warranty of any kind and is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Printed in the U.S.A.

HP StorageWorks Continuous Access EVA performance estimator user guide

Contents

Figures

Tables

Preface

About this guide

This guide provides information about:

- Understanding the impact of distance on replication performance
- Manually estimating replication performance for one outstanding write
- Using the Performance Estimator spreadsheet to estimate replication performance for one outstanding write
- Using the output from the manual calculation or the spreadsheet, determine the minimum expected performance in I/O operations per second (IOPS) for a particular I/O size, separation distance, and intersite link technology
- Estimating the ideal maximum throughput requirements for multiple outstanding writes
- Best practices for using the Performance Estimator

Intended audience

This guide is intended for customers who:

- Purchased the HP StorageWorks Continuous Access Enterprise Virtual Array (EVA)
- Want to estimate the effects of distance on applications that use HP StorageWorks Continuous Access EVA

Prerequisites

Prerequisites for using this product include knowledge of:

- SAN fabric configurations
- Disaster planning
- HP StorageWorks Enterprise Virtual Array (EVA)

Related documentation

In addition to this guide, please see other documents for this product:

- *HP StorageWorks Continuous Access EVA administrator guide*
- *HP StorageWorks Continuous Access EVA planning guide*
- *HP StorageWorks SAN design reference guide*
- *HP StorageWorks Continuous Access and Data Replication Manager SAN Extensions Reference Guide*

These and other HP documents can be found on an HP web site: http://www.hp.com/go/SANDesignGuide.

Document conventions and symbols

Table 1 Conventions

Convention	Element
Medium blue text: Figure 1	Cross–reference links and e–mail addresses
Medium blue, underlined text (http://www.hp.com)	Web site addresses
Bold font	• Key names • Text typed into a GUI element, such as into a box • GUI elements that are clicked or selected, such as menu and list items, buttons, and check boxes
Italics font	Text emphasis
`Monospace font`	• File and directory names • System output • Code • Text typed at the command–line
`Monospace, italic font`	• Code variables • Command–line variables
`Monospace, bold font`	Emphasis of file and directory names, system output, code, and text typed at the command–line

△ CAUTION:

Indicates that failure to follow directions could result in damage to equipment or data.

NOTE:

Provides additional information.

HP technical support

Telephone numbers for worldwide technical support are listed on the HP web site: http://www.hp.com/support/.

Collect the following information before calling:

- Technical support registration number (if applicable)
- Product serial numbers
- Product model names and numbers
- Applicable error messages
- Operating system type and revision level
- Detailed, specific questions

For continuous quality improvement, calls may be recorded or monitored.

HP strongly recommends that customers sign up online using the Subscriber's choice web site at http://www.hp.com/go/e-updates.

- Subscribing to this service provides you with email updates on the latest product enhancements, newest versions of drivers, and firmware documentation updates as well as instant access to numerous other product resources.
- After signing up, you can quickly locate your products by selecting **Business support** and then **Storage** under Product Category.

HP-authorized reseller

For the name of your nearest HP-authorized reseller:

- In the United States, call 1–800–282–6672.
- Elsewhere, see the HP web site: http://www.hp.com. Click Contact HP to find locations and telephone numbers.

Helpful web sites

For other product information, see the following web sites:

- http://www.hp.com
- http://www.hp.com/go/storage
- http://www.hp.com/support
- http://www.docs.hp.com

Providing Feedback

We welcome your feedback!

For HP Command View EVA, please mail your comments and suggestions to CVFeedback@hp.com

For HP Business Copy EVA and HP Continuous Access EVA, please mail your comments and suggestions to EVAReplication@hp.com

1 Introduction

Overview

When you are designing an HP StorageWorks Continuous Access solution, it is important to understand and plan the distance between the two sites to maximize performance and maintain optimal disaster tolerance. A shorter distance between sites will improve performance, but it will also increase the risk of a disaster affecting both sites (reduced disaster tolerance). A longer distance between sites will reduce the risk of disaster affecting both sites (improved disaster tolerance), but it will also decrease performance. The key to a successful solution is creating a balance between performance and protection.

This chapter describes the replication process and the effect of bandwidth and distance on replication performance. Using these two components, you can use this guide to estimate:

- The performance capabilities of an intersite link technology.
- The link requirements for a given level of performance.

Replication process

To complete one replication I/O, HP Continuous Access EVA requires one round trip through the intersite link as follows:

1. The source array sends the data to the destination array.
2. The destination array receives and stores the data in its local cache and returns an acknowledgement to the source array.

Bandwidth and replication performance

HP Continuous Access products can move data at extreme distances. However, the inherent propagation delays in data transfers can affect the amount of time it takes to complete replication I/O. At these extreme distances, bandwidth is not typically the limiting factor in replication performance; rather it is the delay imposed by that distance. At short distances, bandwidth affects performance more than the distance–induced delay does.

To better understand the relationship between bandwidth, distance, and performance, consider a low bandwidth intersite link and a high bandwidth intersite link that are both moving I/O packets containing the same amount of data (Figure 1). The packets are moving from site A to site B (left to right), with the leading edge of both packets arriving at site B at the same time. The difference in link size (bandwidth) enables the high bandwidth link to complete delivery of the packet (loading and unloading) before the low bandwidth link.

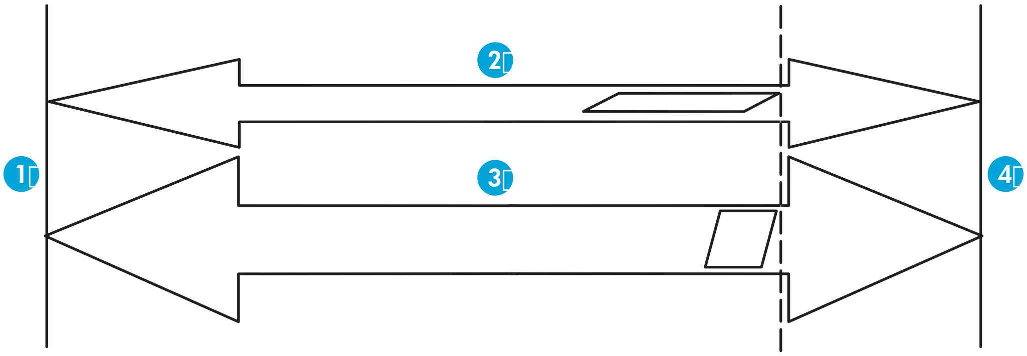


Figure 1 Distance and I/O rate

Callouts:

1. Site A
2. T3 link (44.5 Mb/s)
3. OC3 link (155 Mb/s)
4. Site B

NOTE:

For a visual reference, a nanosecond of fiber is approximately eight inches long. A bit is four inches long at 2 Gb/s and eight inches long at 1 Gb/s. Therefore, a nanosecond of fiber contains two bits at 2 Gb/s and one bit at 1 Gb/s. At 100 Mb/s, a bit is 80 inches and at 10 Mb/s, a single bit is 800 inches.

Distance and replication performance

The speed of light will always limit how long it takes to complete a replication I/O because the data can only travel so fast. In a vacuum, the speed of light is well known at 3 times 10^8 meters per second. However in standard fiber optic cable, light is slowed to only travel at 2 * 10^8 m/sec or 2/3 that of light in a vacuum. Looked at another way, each nano-second of fiber represents an 8 inches length.

When performing I/O to locally direct attached devices, the length of the cable in the I/O performance calculations is not considered because it is so short with respect to the time it takes to complete that I/O. In replication, however, the distance between the two copies of the data can be very large when compared to a single I/O.

For example, a 1 block write completes in 0.25 ms without replication, and after adding synchronous replication at zero distance between the copies, the value becomes 0.35 ms. Add 100 km of cable at 5 micro-seconds per kilometer, and the replication of that 1 block now takes 1.35 ms. (The additional time is due only to the distance between the sites at one round trip). Add another 100 km, and the 0.25 ms write now requires 2.35 ms. Considering only the effect of distance, an sample application performing 4000 synchronous writes per second may now only perform 425 writes per second.

2 Manually estimating performance

Overview

This chapter describes how to manually estimate the time needed to complete one I/O replication across a directly connected (or zero distance) intersite link. It also describes how to add the impact of distance to the calculation.

Using the formula without distance

To calculate performance based on link bandwidth, use the following formula:

```
Y = mX + b
```

Where

- `y` is the time (in milliseconds) to complete one outstanding replication I/O.
- `m` is the *slope* of the line representing the link bandwidth.
- `X` is the size of the transfer in kilobytes.
- `mX` is the is the additional time required to complete a transfer (or data packet) larger than 0.512 kilobytes for a given link bandwidth.
- `b` is the amount of time required to replicate a single 0.512 kilobyte write (replication and conversion overhead). This is the y–*intercept* of the line representing the link bandwidth when `X` is zero.

Use the values in Table 2 for the appropriate intersite link technology to complete this formula.

NOTE:
The values in Table 2 were derived experimentally.

Table 2 Intersite link values

Technology	Bandwidth (Mb/s)	Intercept (b)	Slope (m)
2 Gb/s Fibre Channel	2000	0.3416	0.0268
1 Gb/s Fibre Channel	1000	0.3991	0.0332
1 GbE (Gigabit Ethernet) IP	1000	0.4130	0.0338
OC3 IP	155.5	0.3901	0.0758
E4 IP	139.3	0.3876	0.0818
100 Mb/s IP	100	0.3802	0.1052
T3 IP	44	0.3530	0.2070
E3 IP	34.304	0.3340	0.2666
10 Mb/s IP	10	0.2893	0.8872
E1 IP	2.048	1.1748	4.4434
T1 IP	1.54	1.5557	5.9422

Adding distance

To obtain a more realistic estimate of performance, add the effects of distance, which is measured by the time required for the leading edge of the data to traverse the link. Estimate this time using a speed of 2 x 10^8 meters per second for light in standard fiber–optic cable, which is equal to 5 microseconds per kilometer (8 microseconds per mile). For example, if the intersite link distance is 10,000 km (6,200 mi), the one–way time interval is 50,000 microseconds, or 0.05 seconds. This is approximately 25 times the average rotational latency of a 15,000 rpm disk drive at 0.002 seconds.

Therefore, the actual time to move data from the source array to the destination array is the round–trip distance, plus the time to load and unload the data for the link size. These trips consist of one small control packet and one large data packet and add 10 microseconds of latency (per kilometer) to the one–way intersite link distance to complete each replication write. Based on the distance between the two sites, this latency is added to the previously calculated time to complete a zero distance replication write.

Calculating performance for one outstanding write

During synchronous replication, an application performs one write and waits for that write to complete before issuing the next outstanding write. This section describes how to calculate performance for one outstanding write that contains a 32 KB data packet. Based on the one–way network latency of 1 ms, the local and remote sites are 200 km (125 mi) apart.

Calculating replication time

To estimate the performance of a single 32 KB write using different sized intersite links:

1. Use the values in Table 2 to calculate the loading time for a 32 KB data packet. Multiply the slope by the size of the transfer–m times x). The result of this step is shown in Table 3.

2. Calculate and add the time it takes for a round trip. In this example, the intersite distance is 200 km, therefore the latency is calculated as follows:

   ```
   200 km x 5 ms/km x 2 = 2 ms
   ```

3. Add the replication overhead from Table 2 (the Intercept column).

   The sum is the total time it takes to replicate a 32 KB data packet for a single outstanding write. The result is shown in Table 4.

Calculating time to load

The *time–to–load* value is defined as the length of the data packet in time (the time to load and unload it). Using the intersite link bandwidth and slope values from Table 2, Table 3 shows the time–to–load value (step 1) for a 32 KB data packet using different intersite link technologies.

Table 3 Time–to–load value

Intersite link technology	Slope	Time to load (ms)
2 Gb/s Fibre Channel	0.0268	0.86
1 Gb/s Fibre Channel	0.0332	1.06
1 GbE (Gigabit Ethernet) IP	0.0338	1.08
OC3 IP	0.0758	2.43
E4 IP	0.0818	2.62
100 Mb/s IP	0.1052	3.37
T3 IP	0.2070	6.62
E3 IP	0.2666	8.53
10 Mb/s IP	0.8872	28.39
E1 IP	4.4434	142.19
T1 IP	5.9422	190.15

Calculating time to complete

The *time–to–complete* value is defined as the time to complete one I/O replication across a given technology for a given packet size. It is the time for a host to send a synchronous replication write and receive an acknowledgement of completion. Table 4 show the *time–to–complete* value for a 32

KB data packet (the sum of step 2 and step 3) by adding transfer latency and replication overhead to the results in Table 3.

Table 4 Time–to–complete value

Intersite link technology	Time to load (ms)	+ Transfer latency (ms)	+ Overhead (ms)	= Time to complete I/O (ms)
2 Gb/s Fibre Channel	0.86	2	0.3416	3.20
1 Gb/s Fibre Channel	1.06	2	0.3991	3.46
1 GbE (Gigabit Ethernet) IP	1.08	2	0.4130	3.49
OC3 IP	2.43	2	0.3901	4.82
E4 IP	2.62	2	0.3876	5.01
100 Mb/s IP	3.37	2	0.3802	5.75
T3 IP	6.62	2	0.3530	8.98
E3 IP	8.53	2	0.3340	10.87
10 Mb/s IP	28.39	2	0.2893	30.68
E1 IP	142.19	2	1.1748	145.36
T1 IP	190.15	2	1.5557	193.71

Determining maximum I/Os per second

You can use the time–to–complete value to determine the maximum number of synchronous replication writes that can be completed every second, assuming the next I/O starts immediately after the current I/O completes. To do so, you must invert the the time to complete value into I/Os per second (IOPS). Table 5 shows the results of the inversion for a 32 KB data packet, which were calculated as follows:

- Throughput is the packet size multiplied by the I/O rate (IOPS).
- Bandwidth used is the peak bandwidth of the link, divided by the throughput and then multiplied by 100.

NOTE:

Because most applications produce multiple asynchronous host I/Os, Table 5 shows the minimum expected performance, based on synchronous replication.

Table 5 Bandwidth used for a 32 KB write

Intersite link technology	Approximate IOPS	Throughput (Mb/s)	Approximate single stream bandwidth used
2 Gb/s Fibre Channel	312.58	100.03	5.0%
1 Gb/s Fibre Channel	288.89	92.45	9.2%
1 GbE (Gigabit Ethernet) IP	286.16	91.57	9.2%
OC3 IP	207.65	66.45	42.9%
E4 IP	199.79	63.93	46.0%
100 Mb/s IP	174.02	55.69	55.7%
T3 IP	111.40	35.65	79.2%
E3 IP	92.04	29.45	86.6%
10 Mb/s IP	32.59	10.43	100.0%
E1 IP	6.88	2.20	100.0%
T1 IP	5.16	1.65	100.0%

3 Using the Performance Estimator

Overview

This chapter describes how to use the Performance Estimator, a Microsoft Excel–based spreadsheet, to perform the calculations described in Chapter 2.

Accessing the spreadsheet

The Performance Estimator spreadsheet is available from the following web site:

http://h18006.www1.hp.com/products/storage/software/conaccesseva/index.html

Select *Related information* and then select *HP StorageWorks Continuous Access EVA Performance Estimator* to download the spreadsheet.

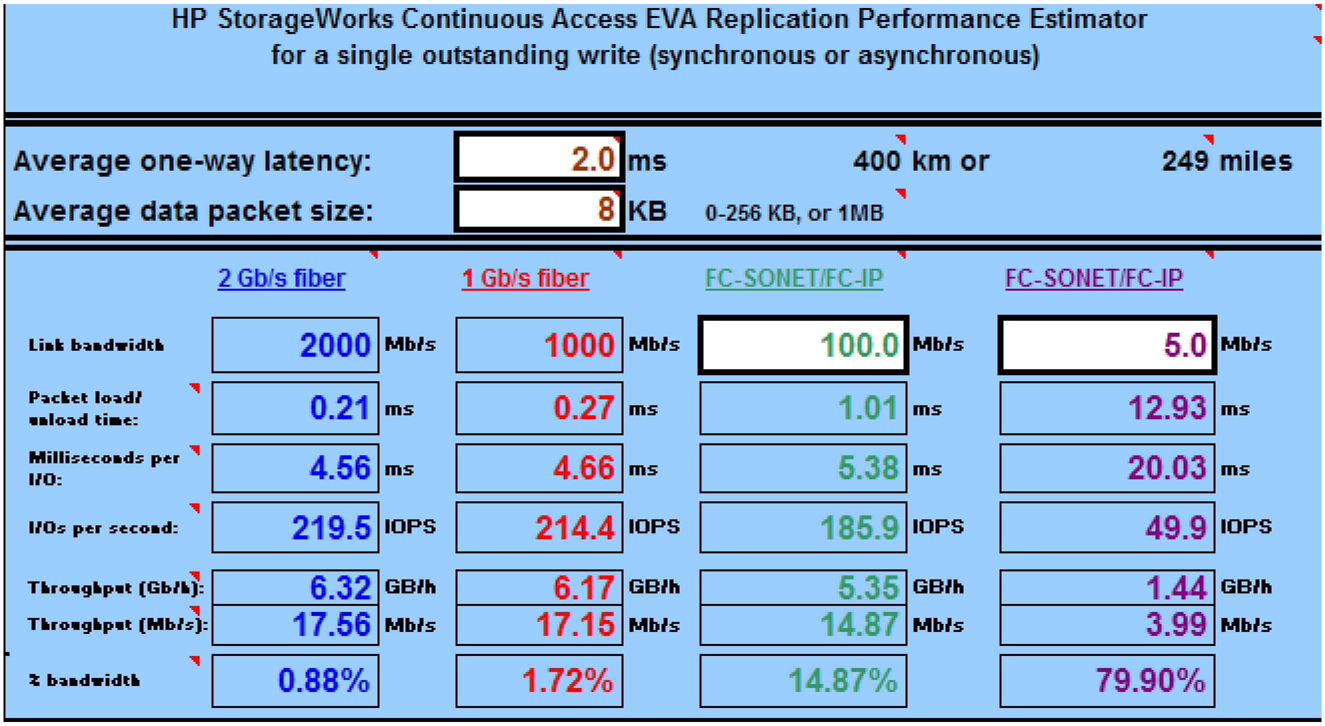
Figure 2 shows the default view of the spreadsheet when you open it. When you enter values in the latency and size fields, the spreadsheet calculates the bandwidth for:

- *2 Gb/s fiber*–Direct fiber or wave dimension multiplexing (WDM) only.
- *1 Gb/s fiber*–Direct fiber or WDM only.
- *FC–SONET or FC–IP*–See Table 2 on for a complete list of the intersite link technologies in this category.

The 2 Gb/s and 1 Gb/s intersite links include long–distance direct fiber connections using longwave or very long distance GBIC/SFP, of either coarse– or dense–wave division multiplexing (CWDM or DWDM).

NOTE:

Place the mouse over any red triangle in the spreadsheet to view additional information for that field.



HP StorageWorks Continuous Access EVA Replication Performance Estimator
for a single outstanding write (synchronous or asynchronous)

Average one-way latency:	2.0	ms	400 km or	249 miles
Average data packet size:	8	KB	0-256 KB, or 1MB	

	2 Gb/s fiber		1 Gb/s fiber		FC-SONET/FC-IP		FC-SONET/FC-IP	
Link bandwidth	2000	Mb/s	1000	Mb/s	100.0	Mb/s	5.0	Mb/s
Packet load/ unload time:	0.21	ms	0.27	ms	1.01	ms	12.93	ms
Milliseconds per I/O:	4.56	ms	4.66	ms	5.38	ms	20.03	ms
I/Os per second:	219.5	IOPS	214.4	IOPS	185.9	IOPS	49.9	IOPS
Throughput (Gb/h):	6.32	GB/h	6.17	GB/h	5.35	GB/h	1.44	GB/h
Throughput (Mb/s):	17.56	Mb/s	17.15	Mb/s	14.87	Mb/s	3.99	Mb/s
% bandwidth	0.88%		1.72%		14.87%		79.90%	

Figure 2 Performance Estimator spreadsheet

Entering one–way latency

To begin using the spreadsheet, enter the appropriate value in the Average one–way latency field. This is the distance between the source and destination arrays in milliseconds. Use the `ping` command to determine round–trip latency and then divide the result in half. If the link is not available, you can also estimate latency using one of the following calculations:

- For point–to–point networks, multiply the driving distance between sites by 1.5.
- For routed networks, multiply the driving distance between sites by 2.25.

When you enter a value and either press *Enter* or click outside the cell, the latency is translated into kilometers and miles. Figure 3 shows three examples.

NOTE:
The maximum value for this field is 100 milliseconds.

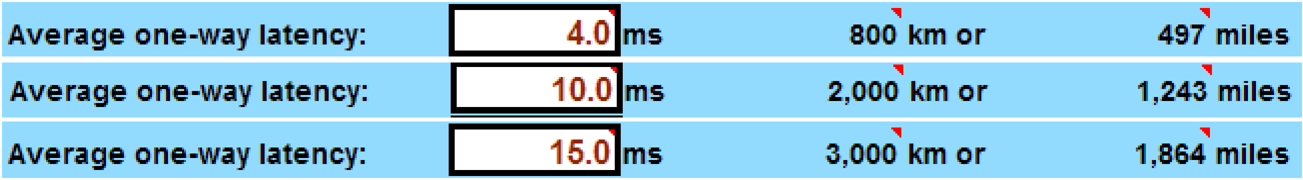
Using the array management server, enter the following command at the command prompt:

```
> ping -n 3600 -l 2048
```

where `-n` is the number of tests at one per second, and `-l` (letter L) is the length of the data packet, a 2048 Byte Fibre Channel frame.

NOTE:

- Check the actual `ping` results with those based on the driving distance. If the estimate based on the driving distance is less than half of the actual value, then ask the network vendor to explain the routing and the reason for the difference.
- If the intersite network uses two distinct paths with different latencies, start with the higher value to get the upper bound using worse case delay.



Average one-way latency:	4.0	ms	800 km or	497 miles
Average one-way latency:	10.0	ms	2,000 km or	1,243 miles
Average one-way latency:	15.0	ms	3,000 km or	1,864 miles

Figure 3 Latency examples

Entering data packet size

The other input is the average size of the application write being replicated to the destination array. The values for this field are 0–256 KB, or 1 MB for a full copy. You can estimate the size for the type of applications you are using. Some examples are:

- 4 KB (such as Microsoft Exchange)
- 8 or 16 KB (small databases, such as Microsoft SQL)
- 32 or 64 KB (large databases, such as Oracle)

NOTE:

If you have writes of multiple sizes (for multiple applications), complete an estimate for each write size, and add the results to obtain an overall network requirement for worst case operation.

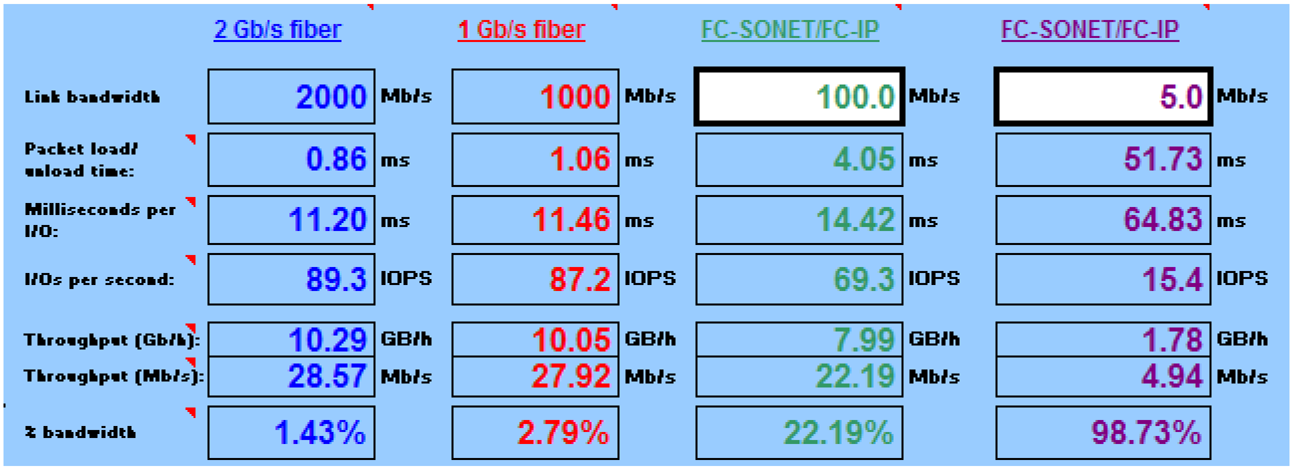
Reading the results

Using the information you enter, the Replication Performance Estimator spreadsheet calculates how latency and write size affect the bandwidth of the applicable intersite link technology.

For all intersite link technologies listed (2 Gb/s fiber, 1 Gb/s fiber, and FC–SONET or FC–IP), the bandwidth results display as follows:

- *Link bandwidth*–The amount of bandwidth (in Mb/s)
- *Packet load/unload time*–The length of time (in milliseconds) to move the data packet on to and off of the intersite link. It is based on the size of the outstanding write and the link bandwidth.
- *ms per I/O*–The length of time to complete one synchronous I/O across the link. Starting with the intersite latency and packet transmit time formulas above. It is the summation of packet transmit time, plus twice the one–way latency, plus all conversion overhead.
- *I/O per second*–The inverse of milliseconds per I/O. The maximum number of synchronous I/O per second for a single data stream. For synchronous I/O, the next I/O does not start until the one in progress has finished. The maximum rate can be achieved only when the next write starts immediately after completion of the previous write. In a real–world environment in which writes are generated in a pseudo–random fashion, the expected peak is approximately 70% of the theoretical peak. The average rate typically does not exceed 50% of the theoretical peak. Table 5 on shows the I/Os per second of various SAN technologies.
- *Throughput (GB/h)*–The transfer rate based on one hour of I/Os per second, multiplied by the data packet size.
- *Throughput (Mb/s)*–The transfer rate based on one second of I/Os per second multiplied by the data packet size. Add these numbers for each I/O stream to get estimated bandwidth requirements.
- *% bandwidth*–The estimated bandwidth required for this single replication. It is based on the estimated throughput and the link bandwidth. Due to the mathematics of the model, the value may exceed 100%.

Figure 4 shows the results for latency of 5 milliseconds and data packet size of 32 KB.



	2 Gb/s fiber		1 Gb/s fiber		FC-SONET/FC-IP		FC-SONET/FC-IP	
Link bandwidth	2000	Mb/s	1000	Mb/s	100.0	Mb/s	5.0	Mb/s
Packet load/unload time:	0.86	ms	1.06	ms	4.05	ms	51.73	ms
Milliseconds per I/O:	11.20	ms	11.46	ms	14.42	ms	64.83	ms
I/Os per second:	89.3	IOPS	87.2	IOPS	69.3	IOPS	15.4	IOPS
Throughput (Gb/h):	10.29	GB/h	10.05	GB/h	7.99	GB/h	1.78	GB/h
Throughput (Mb/s):	28.57	Mb/s	27.92	Mb/s	22.19	Mb/s	4.94	Mb/s
% bandwidth	1.43%		2.79%		22.19%		98.73%	

Figure 4 Sample results

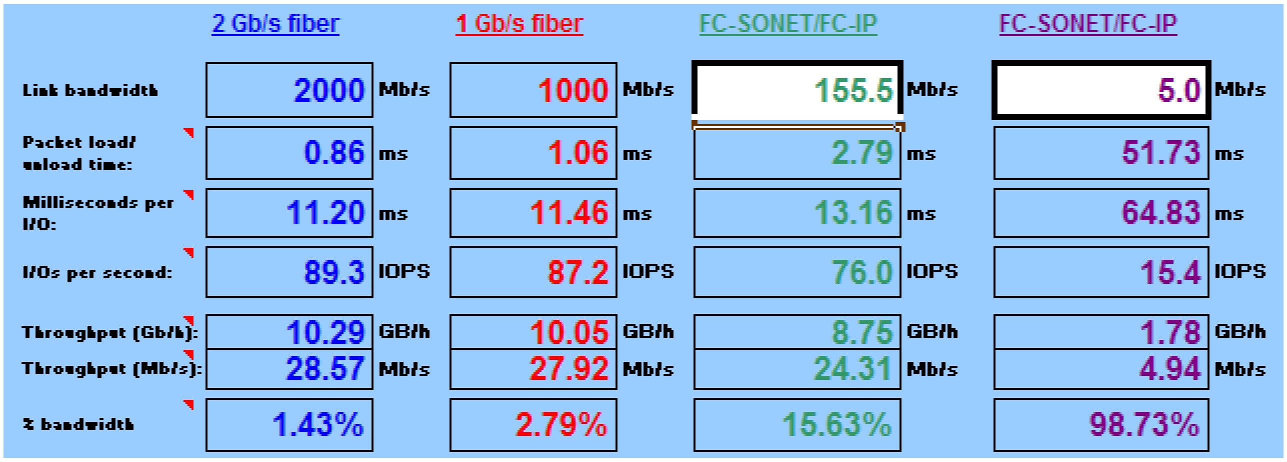
Altering the link bandwidth

You cannot change the link bandwidth fields for 2 Gb/s and 1 Gb/s fiber. You can, however, alter the link bandwidth for both FC–SONET/FC–IP fields. This enables you to enter a value for the specific intersite link technologies you are using. The default values are 100 Mb/s (100 Mb/s Ethernet IP ISL) and 10 Mb/s (the minimum supported bandwidth for each ISL for a single ISL connection). The minimum supported bandwidth for a dual link configuration is 2 Mb/s (E1). Figure 4 shows the results based on these default values. See the *HP Storageworks continuous access and data replication manager SAN extensions reference guide* to see the parameters for a given switch and gateway pair.

If you want to see the results for the OC3 IP intersite link, enter `155.5` in the Link bandwidth field for the first FC–SONET/FC–IP column. Figure 5 shows the results.

NOTE:

The comment for the FC–SONET/FC–IP column lists the value to enter for the corresponding ISL.



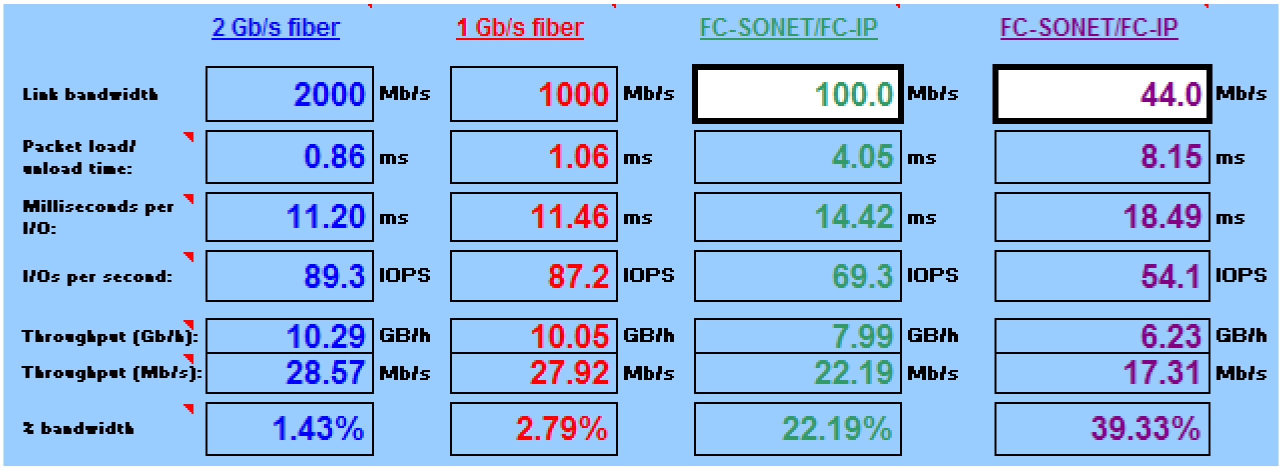
	2 Gb/s fiber	1 Gb/s fiber	FC-SONET/FC-IP	FC-SONET/FC-IP
Link bandwidth	2000 Mb/s	1000 Mb/s	155.5 Mb/s	5.0 Mb/s
Packet load/ unload time:	0.86 ms	1.06 ms	2.79 ms	51.73 ms
Milliseconds per I/O:	11.20 ms	11.46 ms	13.16 ms	64.83 ms
I/Os per second:	89.3 IOPS	87.2 IOPS	76.0 IOPS	15.4 IOPS
Throughput (Gb/h):	10.29 GB/h	10.05 GB/h	8.75 GB/h	1.78 GB/h
Throughput (Mb/s):	28.57 Mb/s	27.92 Mb/s	24.31 Mb/s	4.94 Mb/s
% bandwidth	1.43%	2.79%	15.63%	98.73%

Figure 5 Sample results for OC3 IP intersite link

As another example, you want to see the bandwidth results for the T3 IP intersite link. Change the Link bandwidth value for the first FC–SONET/FC–IP column to the default (100) and enter `44` in the second FC–SONET/FC–IP column. Figure 6 shows the results.

NOTE:

HP recommends that you begin with the default values for the FC–SONET/FC–IP fields. As your analysis proceeds, adjust one value to compare with the default value.



	2 Gb/s fiber	1 Gb/s fiber	FC-SONET/FC-IP	FC-SONET/FC-IP
Link bandwidth	2000 Mb/s	1000 Mb/s	100.0 Mb/s	44.0 Mb/s
Packet load/ unload time:	0.86 ms	1.06 ms	4.05 ms	8.15 ms
Milliseconds per I/O:	11.20 ms	11.46 ms	14.42 ms	18.49 ms
I/Os per second:	89.3 IOPS	87.2 IOPS	69.3 IOPS	54.1 IOPS
Throughput (Gb/h):	10.29 GB/h	10.05 GB/h	7.99 GB/h	6.23 GB/h
Throughput (Mb/s):	28.57 Mb/s	27.92 Mb/s	22.19 Mb/s	17.31 Mb/s
% bandwidth	1.43%	2.79%	22.19%	39.33%

Figure 6 Sample results for T3 IP intersite link

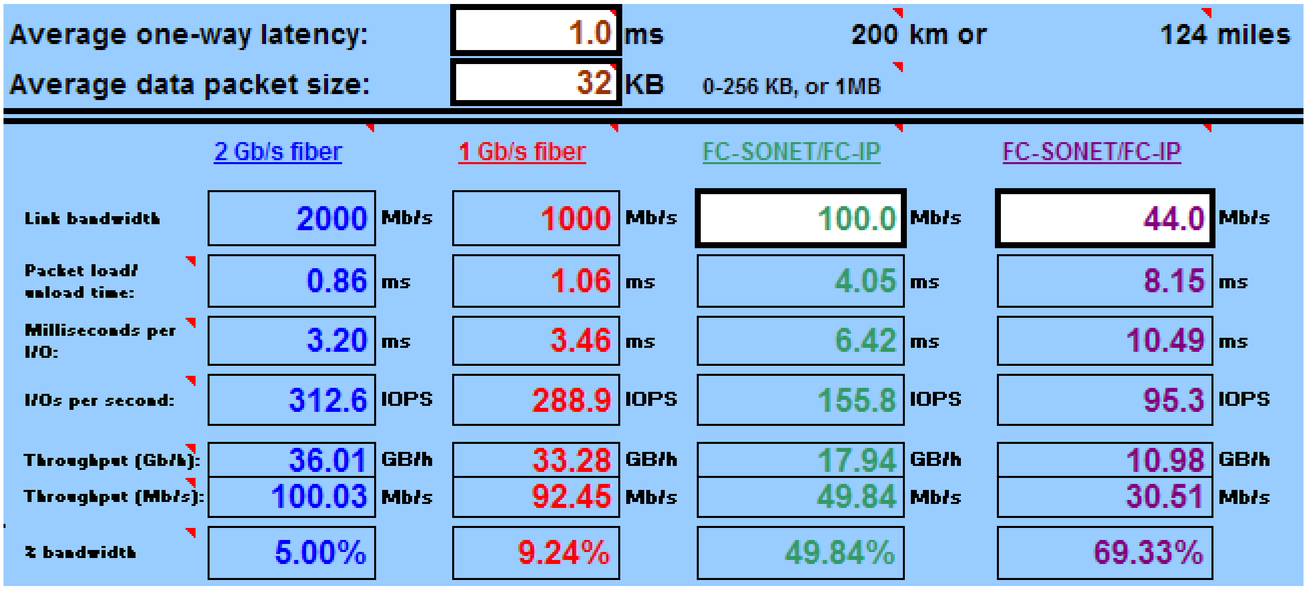
Comparing calculation results

Figure 7 shows that latency of 1 millisecond and data packet size of 32 KB consumes approximately 50% of a 100 Mb/s FC–IP intersite link. The average load on any link must not exceed 40% of its rated capacity, and the peak loading must not exceed 45% of rated capacity. This limitation allows I/O from a failed link or fabric to run on the nonfailed fabric or link without causing additional failures by overloading the active fabric.

The result of the manual calculation for the same latency and data packet size is 55.77% (See Table 3).

NOTE:

There may be a difference between the results you generate manually and those you generate using the spreadsheet. The Performance Estimator spreadsheet uses a mathematical model based on the IP link bandwidth to estimate the slope and intercept of the line that is used in the first two rows of calculations.



Average one-way latency: 1.0 ms 200 km or 124 miles

Average data packet size: 32 KB 0-256 KB, or 1MB

	2 Gb/s fiber		1 Gb/s fiber		FC-SONET/FC-IP		FC-SONET/FC-IP	
Link bandwidth	2000	Mb/s	1000	Mb/s	100.0	Mb/s	44.0	Mb/s
Packet load/ unload time:	0.86	ms	1.06	ms	4.05	ms	8.15	ms
Milliseconds per I/O:	3.20	ms	3.46	ms	6.42	ms	10.49	ms
I/Os per second:	312.6	IOPS	288.9	IOPS	155.8	IOPS	95.3	IOPS
Throughput (Gb/h):	36.01	GB/h	33.28	GB/h	17.94	GB/h	10.98	GB/h
Throughput (Mb/s):	100.03	Mb/s	92.45	Mb/s	49.84	Mb/s	30.51	Mb/s
% bandwidth	5.00%		9.24%		49.84%		69.33%	

Figure 7 Results for a 32 KB write

4 Determining maximum bandwidth

Overview

The previous chapters showed results for one outstanding write. However, most applications do not issue one write and wait for it to complete before immediately sending the next one. Instead, multiple writes are replicated simultaneously (asynchronous replication), issued from different application threads for asynchronous host I/O. This requires that you determine the effect of multiple writes from a single application. This chapter describes how to perform those calculations.

Maximum number of messages

A high bandwidth link can accommodate more data than a low bandwidth link (Figure 8). Also, a longer link can accommodate more data than a shorter link. The maximum capacity is called the *bandwidth–latency product*. To calculate the maximum number of simultaneous messages allowed in the communications link:

1. Multiply the net performance of the communications link (the total bits per second, minus overhead) by the one–way intersite latency (in seconds).
2. Convert the result into bytes (Fibre Channel uses 10 bits per byte) and divide by the average message size (in bytes).

In calculating this number, it is assumed that the application can issue I/O as soon as space is available on the link, thus streaming the data from the host to the source array. This data is then replicated to the destination EVA with acknowledgment back to the host. It is also assumed that because there is only one source, data is written regularly and consistently. These assumptions are necessary to understand peak performance. This calculation is the same for synchronous or asynchronous replication, as either method must move data from the source to the destination.

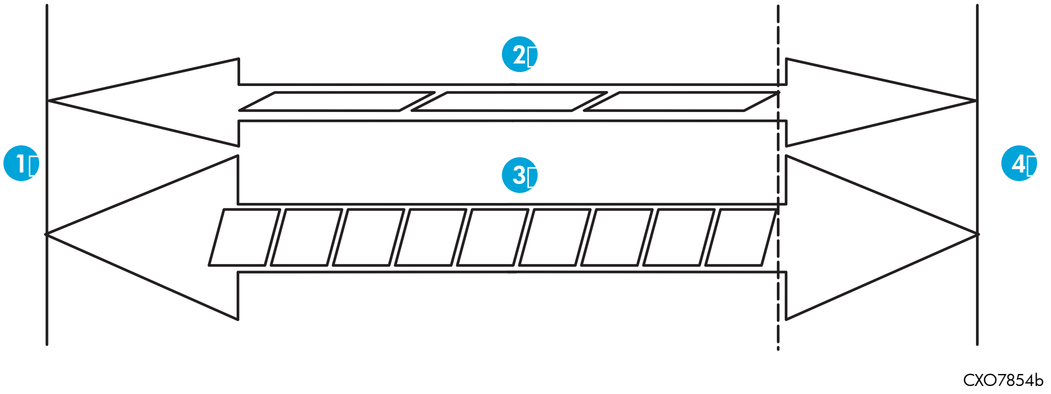


Figure 8 Multiple writes for a single application

Callouts:

1. Site A
2. T3 link
3. OC3 link

4. Site B

Peak write rate

In Figure 8, the parallelograms represent multiple writes of data from the same application. While each packet in the diagram is the same size, representing the same amount of data, they do not have to be of equal size, instead use the average for all replicated writes. The low bandwidth link can accommodate three packets, but the high bandwidth link can accommodate nine packets. It is assumed that each new packet immediately follows the preceding packet, regardless of link bandwidth.

To determine the approximate peak write rate for an application:

1. Invert the single write percentage of bandwidth used (Table 5) and divide by 100. The result is the maximum number of data streams that can be supported by that link for that distance.
2. Multiply the result by the single write rate.

Limitations

Between any two array controllers in a replication relationship, a bidirectional tunnel is created to transport the data (full copies and outstanding writes). A tunnel is a path that exists within the intersite link between two controllers and supports independent data transfers. With two controllers in each array, the four controllers create four tunnels (two active and two standby). A DR group belongs to only one tunnel, but one tunnel can support multiple DR groups.

The number of replication relationships required determines the total number of tunnels that are created (four for each relationship). Further, the number of tunnels determines the minimum required bandwidth for a specific communications link. For example, if two source arrays share the same destination array (relationship fan-in), and do so sharing the same communications link, then eight tunnels are created (four active and four standby). Add the throughput requirements for each active tunnel using a given intersite link technology. This sum (the total throughput requirements for all tunnels) determines the maximum required bandwidth for that communications link.

The following factors can limit the expected maximum number of writes in the link:

- *Raw capability of the host, HBA, and the source array controller*
  For example, to ensure write order across all members of a DR group, the members are restricted to using the same HBA port and array controller and Secure Path's load balancing is turned off.
- *Number of buffer–to–buffer credits allowed between two Fibre Channel devices*
  For each replication relationship, buffers process the replication I/Os. The number of buffers depends on the intersite latency. For example, if the intersite latency is small, not all buffers are used. As the intersite latency increases, the number of buffers used will increase, which may become a performance limitation.

  Table 6, Table 7, and Table 8 list the buffer–to–buffer credits for B–series, C–series, and M–series switches, respectively. This limit becomes the bottleneck on a long–distance direct Fibre Channel connection with very long–distance gigabit interface converters (GBICs) or a wavelength division multiplexing (WDM) solution. It is not usually seen in HP Continuous Access EVA over IP configurations because the credit is returned by the IP gateway to the sending switch.

Replication consumes as much as 492 KB (62 x 8 KB), but the full copy consumes as much as 1 MB (8 x 128 KB). Combined, these two processes may overwhelm an intersite link if it does not have enough bandwidth. The following two factors explain this in more detail:

- *Maximum number of outstanding writes the array controller can support*
  Using Virtual Controller Software version 3.00 or 3.01, the tunnel within the array controller allows up to 31 (8 KB) outstanding writes per controller port. If the bandwidth latency product of the link can support more than 31 writes (such as in high speed, very long distance HP Continuous Access EVA over IP configurations), the maximum number of 8 KB or smaller outstanding writes in the link will be 31 for each array controller sharing the link.

  Using Virtual Controller Software version 3.02 or later, up to 62 outstanding writes are allowed for each controller port. In a worst–case situation in which one port of each controller uses the same link, the combined limit is 124 (62 for each port) outstanding writes, subject to other limits.

- *Impact of the initial full copy process on the intersite link*
  In all 3.x versions of Virtual Controller Software, the full copy process allocates eight 128 KB data buffers. These buffers are used during the initial full copy (a sequential read/write process in which a complete copy of the source array's content is made on the destination array) or after the write history log becomes full. When the log is full and the DR group members are marked for full copy, either a full copy or a fast resynchronization occurs. Fast resynchronization is the process of only moving those blocks of data that changed instead of the entire contents of the DR group members. This fast resynchronization process uses the same 128 KB full copy buffers to move data that must be copied to the destination array.

VCS V3.025 and later

Starting with VCS 3.025, the amount of data inserted into the link may be throttled due to limited bandwidth in the link. This process is dynamic and varies from a minimum of 12, 8KB replication buffers and 1, 128KB copy buffer up to the full 62, 8KB and 8, 128KB buffers. Because replication and full copy are independent processes, if the full copy is complete, there is usually more bandwidth for replication, but not more than 62 buffers. At minimum supported bandwidths, the bandwidth limits the maximums to about 90% of the bandwidth-latency product, rather than using all possible buffers. For more information, see the *HP Storageworks continuous access and data replication manager SAN extensions reference guide*.

Table 6 FC buffer–to–buffer credits for B–series switches

Switch family	Default credits	Credits available with Extended Fabric License
3xxx at 1 Gb/s	16	60
4xxx at 1 Gb/s	27	60
3xxx at 2 Gb/s	27	64
4xxx at 2 Gb/s	27	64

Table 7 FC buffer–to–buffer credits for C–series switches

Switch family	Default credits	Maximum available credits
All 16–port modules, Fx mode	16	255
E or TE modes	255	255
All 32–port modules	12	N/A

Table 8 FC buffer–to–buffer credits for M–series switches

Switch family	Default credits	Maximum available credits
3xxx	60	60
4xxx	162	162
6xxx	60	60

Effects of multiple writes

Figure 9 represents multiple writes from multiple applications. The difference between Figure 9 and the single application in Figure 8 is that the space between the writes is wider for multiple applications. The wider space reduces the number of writes in the intersite link and also reduces the maximum utilization rate of that link. Any differences in the number of writes in the link are based on how the bandwidth is shared between the applications.

NOTE:

If you are using multiple high–performance applications, HP recommends you add additional pairs of EVA storage arrays to better distribute the work. Additional arrays also improve the performance of HP Continuous Access EVA. In some cases, you will need to add additional intersite links.

In real–world application environments, it is not possible for all writes to arrive at precisely the expected times. Independent applications do not coordinate with each other when sending reads or writes to the source array. This lack of coordination creates space between any two writes on the intersite link. Mathematical queue theory suggests that the expected peak utilization is approximately 70% and the expected average is approximately 50%. In addition, there must be sufficient bandwidth on both fabrics for all traffic, if one fabric fails. If there is an even split between the two fabrics, the 50% rate becomes 25%.

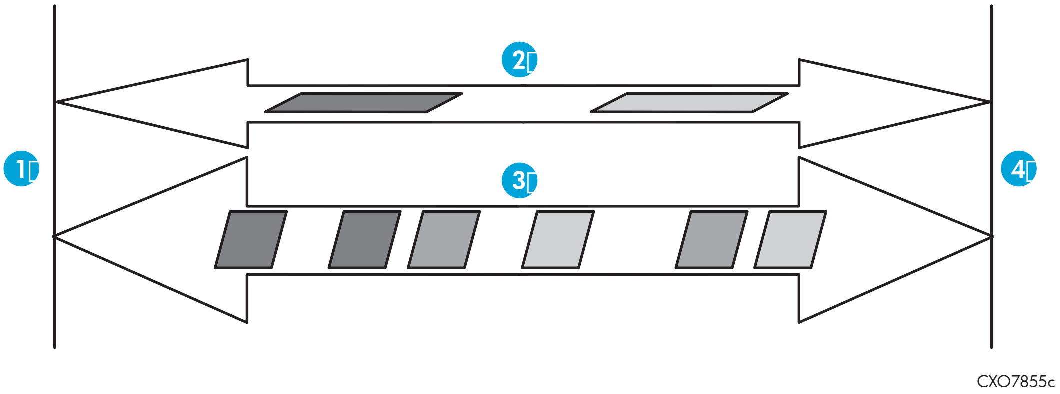


Figure 9 Multiple writes for multiple applications

Callouts:

1. Site A
2. T3 link (44.5 Mb/s)
3. OC3 link (155 Mb/s)
4. Site B

To determine the effect of multiple writes on the intersite link, consider the following example. There are two applications–one performs 32 KB writes and one performs 2 KB writes. Table 5 shows the bandwidth used for a 32 KB write. Table 9, using the same formula as in Table 5, shows the bandwidth used for a 2 KB write.

Table 9 Bandwidth used for a 2 KB write

Intersite link technology	Approximate IOPS	Throughput (Mb/s)	Approximate single stream bandwidth used
2 Gb/s Fibre Channel	417.50	8.35	0.4%
1 Gb/s Fibre Channel	405.60	8.11	0.8%
1 GbE (Gigabit Ethernet) IP	403.13	8.06	0.8%
OC3 IP	393.44	7.87	5.1%
E4 IP	391.97	7.84	5.6%
100 Mb/s IP	386.01	7.72	7.7%
T3 IP	361.40	7.23	16.1%
E3 IP	348.77	6.98	20.5%
10 Mb/s IP	246.08	4.92	49.2%
E1 IP	82.91	1.66	82.9%
T1 IP	64.77	1.30	86.4%

Actual I/Os per second

Determine the actual number of IOPS expected from each application by scaling the single write rate (either up or down) to the expected rate, and then scaling the bandwidth needed by the same number. If the total required bandwidth exceeds 25%, you cannot expect that link to support the replication requirements. For example, one application read and write produces transactions at 3 times the 32 KB single stream rates. The other read and write produces transactions at 4 times the 2 KB single stream rates.

Table 10 shows that only the 2 Gb/s link can support the replication requirements. If high–speed links are not available, you can use multiple low–speed links to obtain the required 25% of available bandwidth.

Table 10 Total bandwidth required for multiple I/O streams

Intersite link technology	32 KB IOPS (x3)	32 KB throughput (Mb/s)	Bandwidth used	2 KB IOPS (x4)	2 KB throughput (Mb/s)	Bandwidth used	Total bandwidth required
2 Gb/s Fibre Channel	938	300	15%	1670	33	2%	17%
1 Gb/s Fibre Channel	867	277	28%	1623	32	3%	31%
1 GbE (Gigabit Ethernet) IP	859	274	27%	1613	32	3%	31%
OC3 IP	623	199	129%	1574	31	20%	149%
E4 IP	599	192	138%	1568	31	23%	161%
100 Mb/s IP	522	167	167%	1544	31	31%	198%
T3 IP	334	107	238%	1446	29	64%	302.%
E3 IP	276	88	260%	1395	28	82%	342%
10 Mb/s IP	98	31	313%	984	20	197%	510%
E1 IP	21	6.6	330%	332	6.6	332%	662%
T1 IP	15	5	330%	259	5.2	345%	676%

△ CAUTION:

Design a solution that will not overload a controller or intersite link during normal operations to prevent significant loss of performance during failure conditions. This is especially true when using asynchronous replication and during the initial full copy or normalization process that occurs when you create a new DR group.

5 Best practices

Overview

This chapter describes how to gather data from an existing application. It also describes network considerations as well as compression and low bandwidth considerations that can affect performance.

Calculating throughput for an existing application

For an application already in production, it is possible to use real data rather than estimates. If more than one application shares the replication link, collect the estimates for each application and add them together. To calculate throughput for an existing application:

1. Capture the peak workloads for a given period of time. Use a tool like PERFMON for Windows or a similar operating system dependent tool to capture the current performance requirements without HP Continuous Access EVA. As a minimum, capture reads per second (IOPS), read throughput per second (MBytes per second), writes per second (IOPS) and write throughput per second (MBytes per second. If possible, collect read and write latency. Do this by application, capturing the data for each physical (Vdisk) as well as any logical disks, if used.
2. Once the data is collected, create a graph of each data set (time on the horizontal and rates on the vertical) so that you can see the peaks during the day. This is critical to understanding whether the peaks occur at the same time. The graph also helps show if the daily average change rate is level or very bursty. If growth is expected, scale these numbers for 12 to 18 months out and use these new numbers as the design goal. Based on this design goal, the peak write rate and throughput will determine how much intersite bandwidth is needed. The combined read and write data shows the ratio of the two, which is used to determine if the EVA itself can support the requirements.
3. When collecting the data, determine the recovery point objective (RPO) and recovery time objective (RTO). RPO is a measure of how much data you can lose due to a problem and tells an architect how real-time the solution needs to be. RTO describes the ideal time to get the recovery site going and usually includes data and application failover and restart. For HP Continuous Access EVA, the design space is an RPO of near zero, with a suggested RTO of several minutes to one or two hours.
4. If you are using IP or SONET as the intersite link, see the*HP Storageworks continuous access and data replication manager SAN extensions reference guide* at:
   http://www.hp.com/go/SANDesignGuide
   to understand the minimum and maximum supported transmission rates for a given FC to IP or FC to SONET gateway and other network requirements.

Other network considerations

The link defined above is minimum required to move the data, but does not consider two other factors: distance and network utilization.

To understand the impact of distance, use the average write size obtained above, and the expected one way latency and enter both into the performance estimator. The numbers will not match because the estimator is based on one outstanding I/O at a time. Chapter 4 describes of the numbers of buffers available for a given type of transfer (up to 62, 8KB writes, or up to 8, 128 KB copies). Multiply the one outstanding results value in the estimator by the number of outstanding writes possible.

NOTE:

If the number is greater than 4 times the zero distance rate for the same data size, consider that the controller performing this replication will be nearly saturated and unable to perform much more work.

Another key factor is utilization due to how the application might create new writes. For example, while SQL tolerates increased write delays due to distance, exchange is very sensitive to increased write delays. For SQL, consider the difference in the required bandwidth based on the peak and that based on the 75th, 85th, and 95th percentile requirements. Because exchange is very sensitive to write delays, consider adding 2x to 4x additional bandwidth to reduce the utilization of the link by the application.

Using compression

All the calculations have assumed that any compression in the system has been turned off. This is done deliberately because not all data is equally compressible. You can use the ZIP utility to reveal how compressible the data is.

NOTE:

Compression causes additional overhead, so if the average data is less than 5% compressible do not use compression. The additional overhead in time (delay) reduces any gains from compression.

Low bandwidth considerations

VCS is designed to detect delays in transmitting the data due to congestion of the low bandwidth link. When these delays are detected, the firmware will first reduce the amount of data sent across the link from the maximum of 62, 8KB buffers and 8, 128KB buffers to minimums of 12, 8KB and 1, 128KB at the minimum supported bandwidth of 2.048 Mbps. At these minimums, the link is full, but unable to support the peak requirements if all of the buffers been used.

NOTE:

If the congestion is only temporary, then the flow control mechanism will increase the numbers of buffers allowed, up to the maximum values.

NOTE:

When using low bandwidth, perform the initial full copy of the data during non-peak hours, when there is no other I/O. In this situation, it is more likely that all 62, 8KB buffers may be used to perform replication.

Estimating initial copy time

Use the performance estimator tool to produce the maximum elapsed time needed to perform the initial copy of the data for a given distance and bandwidth. Enter 1000 into the size field and read the row labeled `throughput (GB/hr)`.

Divide the size of the data to be copied by the rate to get the time value.

Index